УДК: 004.93 https://doi.org/10.15407/jai2025.01.121

П. М. Грицюк[1], М. С. Гаврилюк[2]
[1,2]Національний університет водного господарства та природокористування, Україна
вул. Соборна, 11, м. Рівне, 33028
[1]p.m.hrytsiuk@nuwm.edu.ua
[2]m.s.havryliuk@nuwm.edu.ua
[1]http://orcid.org/0000-0002-3683-4766
[2]http://orcid.org/0000-0003-1149-6251

МОДЕЛЮВАННЯ ВПЛИВУ КЛІМАТИЧНИХ ФАКТОРІВ НА ВРОЖАЙНІСТЬ ПШЕНИЦІ МЕТОДАМИ МАШИННОГО НАВЧАННЯ

Анотація. На сучасному етапі при побудові моделей прогнозування використовуються сучасні методи математичного моделювання, провідну роль у яких відіграють методи машинного навчання та технології штучних нейронних мереж. Кліматичні фактори відіграють вирішальну роль у коливанні врожайності сільськогосподарських культур. Для аналізу їхнього впливу на врожайність пшениці використано методи машинного навчання, зокрема лінійні та нелінійні регресійні моделі. Основною метою дослідження було порівняння лінійних і нелінійних регресійних моделей з різною кількістю параметрів. Враховуючи складні зміни економічних механізмів, які відбулися в українському сільському господарстві за останні тридцять років, ми приходимо до розуміння того, що статистичні дослідження можна проводити лише на даних після 2000 року. Дослідження базується на даних про середньомісячну температуру та кількість опадів у період вегетації пшениці (квітень–червень) за 2000–2021 роки. Дані були згруповані за агрокліматичними зонами, що дозволяє враховувати природно-кліматичні особливості регіонів, а також для підвищення точності моделювання. Результати дослідження підтверджують, що врахування нелінійного впливу кліматичних факторів, таких як квадрати змінних і їхні добутки, значно підвищує точність прогнозування врожайності. Нелінійні моделі продемонстрували майже вдвічі вищу ефективність порівняно з лінійними, що підтверджується значеннями коефіцієнта детермінації (R^2). Збільшення кількості параметрів у моделях також позитивно вплинуло на їхню якість, хоча основну роль відіграли саме нелінійні взаємозв'язки. Отримані моделі дозволяють прогнозувати врожайність пшениці з горизонтом у три місяці, що забезпечує їх практичну цінність для аграрного сектора. Запропонований підхід може бути адаптований для інших сільськогосподарських культур і використаний у різних агрокліматичних зонах, сприяючи підвищенню ефективності управлінських рішень в умовах змін клімату.

Ключові слова: урожайність, кліматичні фактори, машинне навчання.

P. Hrytsiuk[1], M. Havryliuk[2]
[1,2]The National University of Water and Environmental Engineering, Ukraine
11, Soborna st., Rivne, 33028
[1]p.m.hrytsiuk@nuwm.edu.ua
[2]m.s.havryliuk@nuwm.edu.ua
[1]http://orcid.org/0000-0002-3683-4766
[2]http://orcid.org/0000-0003-1149-6251

MODELING THE IMPACT OF CLIMATIC FACTORS ON WHEAT YIELD WITH MACHINE LEARNING

Abstract. At the current stage, modern methods of mathematical modeling are used to build forecasting models, with machine learning and artificial neural network technologies playing a leading role. Climate factors play a crucial role in fluctuations in crop yields. Machine learning methods, including linear and nonlinear regression models, were used to analyze their impact on wheat yields. The main goal of the study was to compare linear and nonlinear regression models with different numbers of parameters. Given the complex changes in economic mechanisms that have taken place in Ukrainian agriculture over the past thirty years, we come to the realization that statistical research can only be conducted on data after 2000. The study is based on the data on average monthly temperature and precipitation during the wheat growing season (April-June) for 2000-2021. The data were grouped by agroclimatic zones, which allows for the consideration of natural and climatic features of the regions, as well as for improving the accuracy of modeling. The results of the study confirm that taking into account the nonlinear influence of climate factors, such as the squares of variables and their products, significantly improves the accuracy of yield forecasting. Nonlinear models demonstrated almost twice the efficiency of linear models, as evidenced by the values of the coefficient of determination (R^2). The increase in the number of parameters in the models also had a positive impact on their quality, although the main role was played by nonlinear relationships. The obtained models allow forecasting wheat yields with a three-month horizon, which

ensures their practical value for the agricultural sector. The proposed approach can be adapted for other crops and used in different agroclimatic zones, contributing to the efficiency of management decisions in the face of climate change.

Keywords: crop yield, climate factors, machine learning.

Вступ

Сільське господарство України відіграє ключову роль у забезпеченні продовольчої безпеки країни та є важливою складовою економіки, що формує значний експортний потенціал. Виробництво зерна, зокрема пшениці, є стратегічним напрямком, що забезпечує стабільний приплив валюти. Середньорічне виробництво зернових за 2019–2021 роки становило близько 75 млн тонн, з яких 50 млн тонн припадало на експорт. У 2021 році Україна досягла рекордного показника – 84 млн тонн зерна. Основною зерновою культурою залишається пшениця, середньорічне виробництво якої становило 26,5 млн тонн за цей період.

Однак, виробництво зернових культур в Україні значною мірою залежить від кліматичних умов, які за останні десятиліття зазнали значних змін. Потепління клімату, нерівномірний розподіл опадів та підвищення середньодекадної температури суттєво вплинули на врожайність. Ці зміни зумовили коригування асортименту вирощуваних культур: зросла частка теплолюбних культур, таких як кукурудза, соя та соняшник, тоді як частка пшениці у загальному врожаї зернових зменшилася з 50% до 40%.

Для забезпечення ефективності сільськогосподарського виробництва в умовах змін клімату особливого значення набуває прогнозування врожайності. Використання методів машинного навчання для моделювання впливу кліматичних факторів відкриває нові можливості для аналізу великих обсягів даних про температуру, опади та інші ключові показники. Це дозволяє створювати точні прогнози, оптимізувати управлінські рішення та мінімізувати ризики.

У цьому дослідженні ми аналізували вплив кліматичних факторів на врожайність пшениці в Україні за період 2000–2021 років. Дані об'єднувалися в регіональні кластери з урахуванням природно-кліматичних характеристик. Такий підхід дозволяє застосовувати сучасні методи математичного моделювання, серед яких провідну роль відіграють алгоритми машинного навчання.

Постановка проблеми

Коливання врожайності пшениці значною мірою залежать від кліматичних факторів, які за останні десятиліття зазнали істотних змін. Потепління клімату, нерівномірний розподіл опадів та підвищення середньої температури впливають на вегетацію рослин та стабільність виробництва. Традиційні методи прогнозування врожайності, які ґрунтуються на лінійних моделях, не завжди враховують складні нелінійні взаємозв'язки між кліматичними факторами та врожайністю. Це знижує точність прогнозів, що унеможливлює ефективне планування аграрного виробництва.

Таким чином, актуальною є потреба у розробці та порівнянні лінійних і нелінійних регресійних моделей, які здатні враховувати як прямий, так і сукупний вплив кліматичних факторів. Важливо визначити, які підходи – додавання нелінійних компонентів чи збільшення кількості параметрів – найбільше покращують точність моделей прогнозування врожайності пшениці.

Аналіз останніх досліджень і публікацій

Прогнозування врожайності є дуже важливим для планування посівів сільськогосподарських культур. Попередня інформація, отримана за допомогою прогнозів, дає змогу всім сторонам, залученим до сільськогосподарського сектору, вжити необхідних заходів, щоб зменшити очікувані збитки. Пшениця, яка є основною продовольчою культурою України, потребує оперативного прогнозування її виробництва для забезпечення внутрішнього споживання та

для підтримки глобальної продовольчої безпеки. Традиційно, розробка емпіричних моделей для прогнозування врожайності ґрунтується на кліматичних даних, супутникових даних або комбінації обох джерел даних. Найбільш популярними методами емпіричного моделювання врожайності є модель регресії та методи машинного навчання. Однак, при цьому, внески різних джерел (клімат, ґрунт, дистанційне зондування) у продуктивність моделі залишаються невідомими. Відсутність чітких порівнянь між продуктивністю підходів на основі моделі регресії та різними методами машинного навчання у прогнозуванні врожайності вимагає подальшого дослідження. Дослідження [1] вирішує цю проблему шляхом об'єднання даних із багатьох джерел для прогнозування врожайності пшениці у провінції Пенджаб у Пакистані. Висновки порівнюються з еталонним тестом, наданим Crop Report Services Punjab, та трьома широко використовуваними методами машинного навчання (метод опорних векторів, метод випадкового лісу та оператор найменшого абсолютного скорочення та вибору LASSO). Для побудови альтернативних емпіричних моделей автори використовували дані про клімат, супутникові дані, характеристики ґрунту та просторову інформацію за період 2017–2022 роки. Отримані результати вказують на те, що алгоритми машинного навчання з використанням об'єднаних наборів даних забезпечують кращу ефективність прогнозування врожайності.

Останні дослідження підтверджують, що кліматичні фактори часто відіграють вирішальну роль у коливанні врожайності сільськогосподарських культур. В першу чергу цей висновок стосується регіонів з континентальним кліматом. З точки зору вирощування пшениці в Україні можна виділити три агрокліматичні зони [2]. У роботі [3] проведено дослідження впливу кліматичних факторів на коливання врожайності пшениці з використанням моделі регресії. Показано, що врахування нелінійного впливу кліматичних факторів значно підвищує точність прогнозу врожайності пшениці. З використанням методики машинного навчання було побудовано моделі для прогнозування майбутньої врожайності з горизонтом прогнозування 3 місяці. Такі прогнози можуть стати основою для прийняття оптимальних інвестиційних та маркетингових рішень на ринку зерна.

У дослідженні [4] було проаналізовано потенційний вплив екстремальних кліматичних умов на виробництво пшениці та бавовни в Південному Пенджабі (Пакистан), використовуючи дані спостережень за 30 років. Сільськогосподарські дані за 1985–2015 роки містили посівні площі та врожайність для шести вибраних районів Південного Пенджабу. Авторами було виявлено потенційний зв'язок між прохолодними та теплими днями і ночами з урожайністю в ті роки, коли виробнича площа залишалася незмінною. У деяких районах було виявлено сильні позитивні кореляції між показниками екстремальних температур і врожайністю.

Метою роботи [5] була спроба оцінити вплив зміни клімату на врожайність пшениці, кукурудзи та соняшнику в Румунії за період 2017–2021 рр., використовуючи статистичні кліматичні дані. Методом порівняння оцінено відхилення зареєстрованих температур повітря та опадів від кліматологічної норми 1981–2010 рр. Графічним методом відображено динаміку місячних температур повітря та кількості опадів у кожному році. Описова статистика для середнього, стандартного відхилення та коефіцієнта варіації забезпечила більш повне відображення змін температури повітря, кількості опадів та врожайності. Кореляційний аналіз та рівняння регресії були використані для встановлення сили впливу кліматичних факторів на врожайність.

У дослідженні [6] було проведено аналіз застосування різних статистичних методів, методів машинного та глибокого навчання для прогнозування врожайності пшениці. При цьому вхідними змінними служать погодні умови та споживання поживних речовин рослинами, а

врожайність пшениці в основних провінціях-виробниках вважається цільовою функцією. Результати аналізу показали, що всі моделі досить добре відображають зв'язок між вибраними змінними середовища та врожайністю пшениці. Метод Random Forest посів перше місце за точністю прогнозування врожайності пшениці, далі йдуть методи CNN, Auto–Arima та LSTM.

У дослідженні [7] було проведено чотири незалежні польові експерименти (регіон центрально-південної Італії), у яких були використані алгоритми машинного навчання для прогнозування врожайності з використанням 16-ти змінних: добриво, управління азотом, кліматичні умови та дані дистанційного зондування. Чотири алгоритми машинного навчання були відкалібровані та перевірені, а модель лінійної регресії використовувалася як контроль. Серед цих чотирьох алгоритмів, модель GBM перевершила інші під час калібрування та можливості перенесення. Показано, що норма внесення азоту, сезонні опади та температура є найважливішими характеристиками для прогнозування врожайності пшениці.

Дослідження [8] було проведене в Канаді та має на меті оцінити потенціал п'яти моделей машинного навчання (Adaptive Boosting (AdaBoost), Gradient Boosting Machine (GBM), XGBoost, LightGBM і Random Forest (RF)) у прогнозуванні врожайності сільськогосподарських культур. Дослідження об'єднало різні фактори, включаючи змінні клімату, супутникові індекси рослинності, характеристики ґрунту та дані перепису медоносних бджіл. Точність моделей оцінювалися за допомогою середньоквадратичної помилки (RMSE), показника R-квадрат та середньої абсолютної помилки (MAE). Дослідження також оцінювало зміни врожайності за різних сценаріїв зміни клімату. Було виявлено мінімальні варіації між сценаріями та значний негативний вплив змін клімату на врожайність ріпаку та сої по всій Канаді. Крім того, автори запропонували стратегію готовності до зміни клімату, допомагаючи фермерам оптимізувати розподіл ресурсів і ефективно керувати ризиками.

Мета дослідження

Метою даного дослідження є моделювання впливу кліматичних факторів на врожайність пшениці з використанням лінійних і нелінійних регресійних моделей з різною кількістю параметрів; оцінка точності прогнозування, а також виявлення ключових факторів, що визначають коливання врожайності у різних агрокліматичних зонах України.

Виклад основного матеріалу

Дослідження статистики врожайності пшениці в областях України за останні 22 роки виявило чітку тенденцію до зростання врожайності [3]. Однак тенденція підвищення врожайності супроводжується значними коливаннями, зумовленими впливом кліматичних факторів (рис. 1).

Основна вегетація пшениці відбувається протягом трьох місяців: квітня, травня та червня. Основні кліматичні фактори, які впливають на врожайність – це середня температура (°C) та кількість опадів (мм). В ролі основного інструмента дослідження впливу кліматичних факторів на врожайність ми вибрали модель лінійної регресії, яка відображає вплив кліматичних факторів на детрендовану врожайність пшениці. Для моделювання динаміки врожайності ми використовували лінійну трендову модель:

$$trend_t = a_0 + a_1 t \qquad (1)$$

Детрендована врожайність eps_t визначалася за співвідношенням:

$$eps_t = y_t - trend_t \qquad (2)$$

Тут: y_t – врожайність; t – час. Детрендована врожайність є функцією кліматичних факторів. Метою даної статті є дослідження впливу кліматичних факторів на коливання детрендованої врожайності

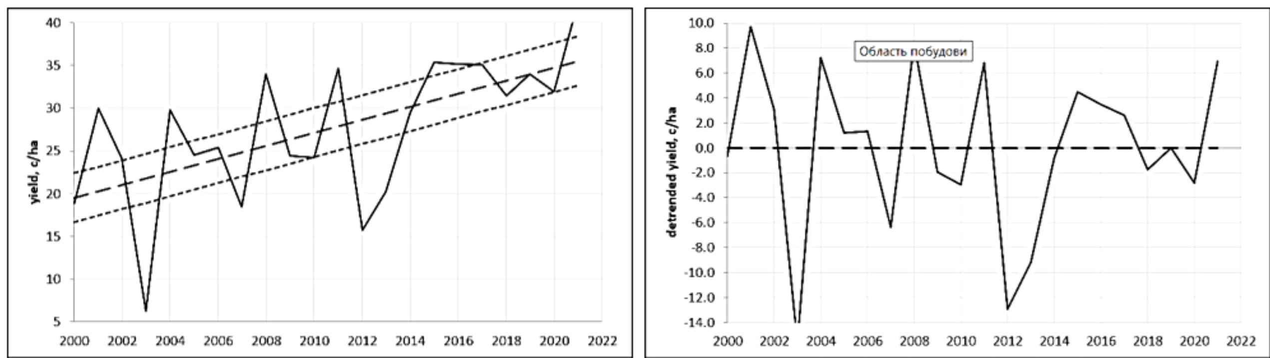


Рис.1. Динаміка врожайності пшениці у Херсонській області (зліва); детрендована врожайність (справа) [3]

з використанням регресійного аналізу та методів машинного навчання.

Схожі природно-кліматичні особливості сусідніх областей України дозволяють згрупувати їх в агрокліматичні зони. Аналіз кореляційної матриці, побудованої на основі часових рядів детрендованої врожайності, дозволив нам виділити три агрокліматичні зони на території України: степова зона (Одеська, Херсонська, Миколаївська, Запорізька, Дніпропетровська та Кіровоградська області); лісостепова зона (Сумська, Харківська, Полтавська, Київська, Черкаська та Вінницька області); західна зона (Волинська, Рівненська, Житомирська, Львівська, Тернопільська та Хмельницька області) [3]. Відмінний характер динаміки врожайності в інших областях України не дозволив включити їх до жодної з вищеназваних зон. Великий обсяг даних по одній агрокліматичній зоні дає можливість використовувати методи машинного навчання [9].

Модель 1. Лінійна модель з 6-ма параметрами

Найпростіший підхід до побудови регресійної моделі передбачає урахування впливу шести кліматичних факторів (три значення середньомісячної температури та три значення місячної суми опадів) на детрендовану врожайність пшениці

$$eps = \alpha_0 + \alpha_4 t_4 + \alpha_5 t_5 + \alpha_6 t_6 + \\ + \beta_4 R_4 + \beta_5 R_5 + \beta_6 R_6 + \varepsilon \qquad (3)$$

Тут t_4, t_5, t_6 – середньомісячні значення температури квітня, травня та червня відповідно; R_4, R_5, R_6 – сума опадів за кожен із вказаних місяців; $\alpha_0, \alpha_4, \alpha_5, \alpha_6, \beta_4, \beta_5, \beta_6$ – коефіцієнти регресії; ε – випадковий залишок моделі. Для побудови лінійної регресійної моделі врожайності типу (3) для степової зони ми використовували 132 зразки, кожен з яких містив 6 кліматичних факторів та значення детрендованої врожайності. У середовищі Python модель реалізується з використанням бібліотек scikit learn та statsmodels. Модель OLS (Ordinary Least Squares) дозволяє оцінити статистичну

Таблиця 1. Параметри якості лінійної регресійної моделі з 6-ма факторами

	F-statistic	Prob(F-statistic)	R-squared	Cond. No.
Модель з використанням усіх даних та всіх факторів	14.05	3.57e-12	0.403	1310
З попередньої моделі були вилучені незначущі фактори	19.58	1.43e-12	0.381	814
Параметри моделі, отримані шляхом перехресної крос-валідації з вилученням незначущих факторів	16.04	6.89e-07	0.291	815
Параметри моделі, отримані шляхом перехресної крос-валідації з врахуванням усіх факторів	12.43	2.78e-06	0.280	1369

значущість моделі регресії та статистичну значущість факторів моделі. Такий підхід дозволяє відсівати незначущі фактори, покращуючи, таким чином, якість моделі.

Спочатку ми побудували OLS модель лінійної регресії з 6-ма кліматичними факторами з використанням усіх даних (132 зразки). Описання моделі наведено у першому рядку табл. 1. Для оцінки якості моделі ми використали наступні фактори: F-statistic (оцінка значущості моделі), R-squared (оцінка точності відтворення даних на тестовій вибірці), Cond. No. (оцінка мультиколінеарності моделі).

Для покращення якості моделі видаляємо незначущі фактори t_6, R_5. Параметри покращеної моделі наведено у другому рядку табл. 1. Значення коефіцієнта детермінації даної моделі дещо знизилося у порівнянні з попередньою моделлю, але інші показники свідчать про покращення якості моделі.

Однак єдина модель, побудована нами вище з використанням усіх даних, не є статистично значущою. Дотримуючись принципів статистичного моделювання [9], ми розділили дані на дві частини: навчальну вибірку (99 зразків, вибрані випадковим чином - 75% даних) і тестову вибірку (33 зразки, не включені в навчальну вибірку – 25% даних). Модель будується на основі даних навчальної вибірки, а потім перевіряється на даних тестової вибірки. Але якість моделі регресії, побудованої з розділенням даних на навчальну та контрольну вибірки, буде сильно залежати від набору даних, які були віднесені до навчальної вибірки. Тому, для підвищення статистичної значущості результатів, доцільно багаторазово розбивати дані на навчальні та тестові набори (перехресна валідація моделі). Після реалізації усіх моделей їх статистичні характеристики усереднюють. У нашому дослідженні ми використовували десятикратну перехресну валідацію моделі з повторенням. Усереднені значення показників якості лінійної регресійної моделі, побудованої шляхом перехресної крос-валідації, наведені у третьому рядку табл. 1.

Середнє значення коефіцієнта детермінації, отримане шляхом перехресної крос-валідації ($R^2 = 0.291$), трохи нижче, ніж відповідний коефіцієнт моделі, побудованої з використанням усіх даних ($R^2 = 0.381$). Однак, такий підхід дозволяє отримати об’єктивну оцінку якості лінійної регресійної моделі. Порівняння двох останніх рядків табл. 1 дозволяє зробити висновок про те, що вилучення незначущих факторів є важливим для покращення якості лінійної регресійної моделі. Цей висновок є статистично значущим, оскільки отриманий при усередненні показників десяти моделей, побудованих у процесі крос-валідації.

Достатньо низьке значення коефіцієнта детермінації побудованих нами вище моделей вимагає нових підходів, спрямованих на підвищення їх якості. Одним з таких підходів є врахування нелінійного впливу кліматичних факторів на врожайність пшениці.

Модель 2. Нелінійна регресійна модель з 6-ма параметрами

Останні дослідження показують важливість врахування нелінійного впливу кліматичних факторів при моделюванні врожайності пшениці [3]. Такий підхід передбачає врахування при побудові регресійної моделі не тільки лінійних факторів $t_4, t_5, t_6, R_4, R_5, R_6$, а також квадратів факторів: $t_4^2, t_5^2, t_6^2, R_4^2, R_5^2, R_6^2$ та добутків послідовних у часі факторів: $t_4 t_5, t_5 t_6, R_4 R_5, R_5 R_6$. Квадрати факторів відображають підвищену чутливість врожайності до погодних і кліматичних умов на окремих етапах вегетації. Добутки послідовних кліматичних факторів відображають сукупний вплив кліматичних факторів на двох послідовних етапах вегетації, який може бути сильнішим від суми впливів окремих факторів.

При дослідженні нелінійної регресійної моделі ми дотримувалися схеми, наведеної вище. Спочатку була побудована модель з використанням усіх даних та усіх впливаючих факторів (перший рядок табл. 2). Після цього була побудована модель на всіх даних, але з вилученням незначущих факторів (другий рядок табл. 2). На останньому етапі була

виконана десятикратна перехресна крос-валідація моделі. Параметри останньої моделі були отримані на контрольних даних (модель з вилученням незначущих факторів – третій рядок табл. 2; модель, побудована з використанням усіх факторів – четвертий рядок табл. 2). Як і у випадку лінійної моделі, вилучення незначущих факторів є важливим для покращення якості регресійної моделі.

Модель вказує на те, що вплив нелінійних факторів є дуже суттєвим для моделювання залежності врожайності

Таблиця 2. Параметри якості нелінійної регресійної моделі

	F-statistic	Prob(F-statistic)	R-squared	Cond. No.
Модель з використанням усіх даних та всіх факторів	12.47	2.18e-18	0.634	311
З попередньої моделі були вилучені незначущі фактори	27.54	1.62e-22	0.609	108
Параметри моделі, отримані шляхом перехресної крос-валідації з вилученням незначущих факторів	21.79	6.44e-10	0.526	120
Параметри моделі, отримані шляхом перехресної крос-валідації з врахуванням усіх факторів	10.72	1.41e-07	0.406	379

пшениці від кліматичних факторів у степовому регіоні. При врахуванні дії нелінійних факторів коефіцієнт детермінації верифікованої регресійної моделі зріс від $R^2 = 0.291$ до $R^2 = 0.526$. Аналізуючи усереднені коефіцієнти нелінійної регресійної моделі з видаленими незначущими факторами (табл. 3), можна зробити висновок, що прохолодний квітень з достатньою кількістю опадів та прохолодний травень є сприятливими для високих врожаїв пшениці у степовій зоні.

Таблиця 3. Усереднені коефіцієнти нелінійної регресійної моделі

const	t_6	R_4	R_6	t_{44}	t_{55}	t_{45}	R_{45}
1.0927	-0.2796	0.1429	0.1611	-3.4978	-2.6791	5.6176	0.1791

Модель 3. Лінійна модель з 12-ма параметрами

Ще одним підходом, спрямованим на підвищення точності прогнозування регресійної моделі, є збільшення кількості кліматичних факторів, які впливають на врожайність пшениці. З цією метою можна використовувати усереднені значення факторів за десятиденний період (перша, друга, третя декада місяця). Однак, такий підхід можна застосувати лише для температури повітря. Для кількості опадів таке подрібнення часового періоду спостережень є недоцільним. Через малу кількість опадів у степовій зоні значення багатьох факторів буть близькими до нуля. Загальний вигляд лінійної регресійної моделі, побудованої з врахуванням 9-ти температурних факторів та 3-х факторів, пов'язаних з опадами, буде наступним

$$eps = \alpha_0 + \alpha_{41}t_{41} + \alpha_{42}t_{42} + \cdots + \alpha_{63}t_{63} + \beta_4 R_4 + \beta_5 R_5 + \beta_6 R_6 + \varepsilon \quad (4)$$

Тут $t_{41}, t_{42}, \cdots, t_{63}$ – середньодекадні значення температури послідовних декад квітня, травня та червня відповідно; R_4, R_5, R_6 – сума опадів за кожен із вказаних місяців; $\alpha_0, \alpha_{41}, \alpha_{42}, \cdots,$ $\alpha_{63}, \beta_4, \beta_5, \beta_6$ – коефіцієнти регресії; ε – випадковий залишок моделі.

Використовуючи, описану вище схему досліджень, ми спочатку побудували OLS модель лінійної регресії з 12-ма кліматичними факторами з використанням усіх даних (132 зразки). Описання моделі наведено у першому рядку табл. 4.

Для покращення якості моделі видаляємо незначущі фактори t_{41}, t_{62}, R_6. Параметри покращеної моделі наведено у другому рядку табл. 4.

Після цього була проведена процедура десятикратної перехресної крос-валідації. Як і раніше, ми розділяли дані випадковим чином на навчальну вибірку (75% даних) і тестову вибірку (25% даних).

Усереднені значення показників якості лінійної регресійної моделі, побудованої шляхом перехресної крос-валідації, наведені у третьому та четвертому рядках табл. 4. Вилучення незначущих факторів слабко впливає на показник *F-statistic*, однак, цей прийом підвищує значення коефіцієнта детермінації. У порівнянні з шестифакторною моделлю лінійної регресії ($R^2 = 0.291$) значення показника *R-squared* 12-факторної моделі збільшилося до $R^2 = 0.340$. Як і у попередніх випадках, вилучення незначущих факторів покращує прогнозну точність лінійної моделі.

Таблиця 4. Параметри якості лінійної регресійної моделі з 12-ма факторами

	F-statistic	Prob (F-statistic)	R-squared	Cond. No.
Модель з використанням усіх даних та всіх факторів	10.46	7.44e-14	0.513	1760
З попередньої моделі були вилучені незначущі фактори	13.78	4.38e-15	0.504	1100
Параметри моделі, отримані шляхом перехресної крос-валідації з вилученням незначущих факторів	12.14	2.09e-06	0.340	1192
Параметри моделі, отримані шляхом перехресної крос-валідації з врахуванням усіх факторів	12.43	2.78e-06	0.280	1369

Модель 4. Нелінійна регресійна модель з 12-ма параметрами

Для врахування нелінійного впливу кліматичних факторів на врожайність пшениці додамо до попередньої моделі квадрати факторів: $t_{41}^2, t_{42}^2, \cdots, t_{63}^2, R_4^2, R_5^2, R_6^2$ та добутки послідовних у часі факторів: $t_{41}t_{42}, t_{42}t_{43}, \cdots, t_{62}t_{63}, R_4R_5, R_5R_6$.

Показники якості нелінійної регресійної моделі з використанням усіх даних та усіх факторів наведені у першому рядку табл. 5. Після цього була побудована модель на всіх даних, але з вилученням незначущих факторів (другий рядок табл. 5). На останньому етапі була виконана десятикратна перехресна крос-валідація моделі. Параметри останньої моделі були отримані на контрольних даних (модель з вилученням незначущих факторів – третій рядок табл. 5; модель, побудована з використанням усіх факторів – четвертий рядок табл. 5).

Таблиця 5. Параметри якості нелінійної регресійної моделі

	F-statistic	Prob(F-statistic)	R-squared	Cond. No.
Модель з використанням усіх даних та всіх факторів	9.21	3.52e-18	0.764	829
З попередньої моделі були вилучені незначущі фактори	23.06	2.92e-27	0.734	373
Параметри моделі, отримані шляхом перехресної крос-валідації з вилученням незначущих факторів	19.00	5.51e-12	0.618	444
Параметри моделі, отримані шляхом перехресної крос-валідації з врахуванням усіх факторів	11.00	1.11e-07	0.013	1374

Регресійна модель з 12-ма вхідними факторами та із врахуванням їх нелінійного впливу на врожайність має найкращі показники якості серед усіх моделей, побудованих нами. У таблиці 6 наведені показники якості різних моделей, верифікованих методом перехресної крос-валідації. В усіх моделях для покращення якості були вилучені незначущі фактори.

Аналіз таблиці 6 дозволяє оцінити ефективність, запропонованих нами підходів, спрямованих на покращення прогнозної точності регресійної моделі залежності врожайності пшениці від кліматичних факторів. Врахування нелінійного впливу кліматичних факторів та збільшення їх кількості дозволяє покращити якість регресійної моделі.

Основну роль у покращенні моделі відіграє врахування впливу нелінійних факторів, а збільшення кількості факторів за рахунок зменшення тривалості часових проміжків спостережень лише незначним чином покращує якість моделі.

Таблиця 6. Параметри якості верифікованих регресійних моделей

	F-statistic	Prob(F-statistic)	R-squared	Cond. No.
Лінійна 6-факторна модель	16.04	6.89e-07	0.291	815
Нелінійна 6-факторна модель	21.79	6.44e-10	0.526	120
Лінійна 12-факторна модель	12.14	2.09e-06	0.340	1192
Нелінійна 12-факторна модель	19.00	5.51e-12	0.618	444

Таблиця 7. Значущі кліматичні фактори, включені до регресійних моделей

	Лінійні фактори	Квадрати факторів	Добутки послідовних факторів
Лінійна 6-факторна модель	t_4, t_5, R_4, R_6		
Нелінійна 6-факторна модель	t_6, R_4, R_6	t_4, t_5	t_4, t_5; R_4, R_5
Лінійна 12-факторна модель	t_{42}, t_{43}, t_{51}, t_{52}, t_{53}, t_{61}, t_{63}, R_4, R_5		
Нелінійна 12-факторна модель	$t_{41}, t_{53}, t_{61}, R_4$	t_{41}, t_{43}, t_{51}, t_{52}, t_{53}, t_{61}, t_{62}	t_{43}, t_{51}; t_{52}, t_{53}; t_{53}, t_{61}

У табл. 7 наведені фактори, включені до описаних у табл. 6 моделей регресії. Видно, що врахування впливу нелінійних факторів є дуже важливим. Кількість нелінійних факторів у другій та четвертій моделях перевищує кількість лінійних факторів. Основну роль у регресійних моделях відіграють кліматичні фактори квітня та травня.

Висновки

Метою цього дослідження було порівняння лінійних і нелінійних регресійних моделей прогнозування врожайності пшениці з різною кількістю параметрів. Результати показали, що врахування нелінійного впливу кліматичних факторів значно підвищує точність прогнозів у порівнянні з лінійними моделями.

Побудовані нелінійні моделі продемонстрували майже подвоєну точність прогнозів порівняно з лінійними завдяки здатності враховувати складні взаємозв'язки між кліматичними факторами та врожайністю. Оптимальну конфігурацію моделей було визначено шляхом варіативного налаштування параметрів і застосування багаторазової k-кратної перехресної перевірки.

Аналіз моделей з різною кількістю параметрів показав, що додавання додаткових змінних до моделей нелінійного типу сприяє суттєвому покращенню їх прогностичних можливостей без надмірного збільшення складності.

Розроблені моделі дозволяють прогнозувати врожайність пшениці на три місяці вперед із високою точністю, що забезпечує їх практичну цінність для фермерів і великих аграрних виробників. Вони також адаптовані до використання в різних агрокліматичних зонах України.

Розроблена нами методика дозволяє краще підготувати аграрний сектор до викликів, пов'язаних зі змінами клімату. Запропонована методика є універсальною та може бути адаптована для моделювання врожайності інших сільськогосподарських культур, що розширює її практичне застосування в аграрній сфері.

Література

1. M. Ashfaq, I. Khan, A. Alzahrani, M. U. Tariq, H. Khan and A. Ghani. (2024) Accurate Wheat Yield Prediction Using Machine Learning and Climate-NDVI

Data Fusion. IEEE Access, Vol. 12, pp. 40947-40961. [Online]. Available: https://doi.org/10.1109/ACCESS.2024.3376735.

2. D. Muller, A. Jungandreas, F. Koch, F. Shirhorn. (2016) The impact of climate change on wheat production in Ukraine. Report on agricultural policy (APD).

3. Petro Hrytsiuk, Maksym Havryliuk. (2024) Modeling of the nonlinear impact of climatic factors on wheat yield using machine learning techniques. 19th International Conference ICTERI-2024, Lviv, 16-21 September 2024.

4. Ali, M.A.; Hassan, M.; Mehmood, M.; Kazmi, D.H.; Chishtie, F.A.; Shahid, I. (2022) The Potential Impact of Climate Extremes on Cotton and Wheat Crops in Southern Punjab, Pakistan. Sustainability, Vol. 14, p. 1609. https://doi.org/10.3390/su14031609.

5. Agatha Popescu, Toma Adrian Dinu, Elena Stoian and Valentin Serban. (2023) Climate Change and its Impact on Wheat, Maize and Sunflower Yield in Romania in the Period 2017-2021. Scientific Papers Series Management, Economic Engineering in Agriculture and Rural Development, Vol. 23(1).

6. Özden C., Karadoğan N. (2024) Wheat Yield Prediction for Turkey Using Statistical Machine Learning and Deep Learning Methods. PAKISTAN JOURNAL OF AGRICULTURAL SCIENCES, Vol. 61(1), pp.1-7.

7. Fiorentini, M., Schillaci, C., Denora, M., Zenobi, S., Deligios, P., Orsini, R., Santilocchi, R., Perniola, M., Montanarella, L., & Ledda, L. (2024). A machine learning modeling framework for Triticum turgidum subsp. durum Desf. yield forecasting in Italy. Agronomy Journal, Vol. 116, pp. 1050–1070. [Online]. Available: https://doi.org/10.1002/agj2.21279.

8. Navid Mahdizadeh Gharakhanlou, Liliana Perez. (2024) From data to harvest: Leveraging ensemble machine learning for enhanced crop yield predictions across Canada amidst climate change. Science of The Total Environment. Vol. 951, p. 175764. [Online]. Available: https://doi.org/10.1016/j.scitotenv.2024.175764.

9. McKinney, W. (2018) Python for Data Analysis. O'Reilly Media.

10. Schierhorn, F., Hofmann, M., Gagalyuk, T. et al. (2021) Machine learning reveals complex effects of climatic means and weather extremes on wheat yields during different plant developmental stages. Climatic Change, Vol. 169(39). [Online]. Available: https://doi.org/10.1007/s10584-021-03272-0.

11. P. Hrytsiuk, T. Babych, S. Baranovsky, M. Havryliuk. (2023) Assessing of Climate Impact on Wheat Yield using Machine Learning Techniques. CEUR Workshop Proceedings, 3513, pp. 314–329.

12. Iqbal N, Shahzad MU, Sherif E-SM, Tariq MU, Rashid J, Le T-V, Ghani A. (2024) Analysis of Wheat-Yield Prediction Using Machine Learning Models under Climate Change Scenarios. Sustainability. Vol. 16(16), p. 6976. [Online]. Available: https://doi.org/10.3390/su16166976.

References

1. M. Ashfaq, I. Khan, A. Alzahrani, M. U. Tariq, H. Khan and A. Ghani. (2024) Accurate Wheat Yield Prediction Using Machine Learning and Climate-NDVI Data Fusion. IEEE Access, Vol. 12, pp. 40947-40961. [Online]. Available: https://doi.org/10.1109/ACCESS.2024.3376735.

2. D. Muller, A. Jungandreas, F. Koch, F. Shirhorn. (2016) The impact of climate change on wheat production in Ukraine. Report on agricultural policy (APD).

3. Petro Hrytsiuk, Maksym Havryliuk. (2024) Modeling of the nonlinear impact of climatic factors on wheat yield using machine learning techniques. 19th International Conference ICTERI-2024, Lviv, 16-21 September 2024.

4. Ali, M.A.; Hassan, M.; Mehmood, M.; Kazmi, D.H.; Chishtie, F.A.; Shahid, I. (2022) The Potential Impact of Climate Extremes on Cotton and Wheat Crops in Southern Punjab, Pakistan. Sustainability, Vol. 14, p. 1609. https://doi.org/10.3390/su14031609.

5. Agatha Popescu, Toma Adrian Dinu, Elena Stoian and Valentin Serban. (2023) Climate Change and its Impact on Wheat, Maize and Sunflower Yield in Romania in the Period 2017-2021. Scientific Papers Series Management, Economic Engineering in Agriculture and Rural Development, Vol. 23(1).

6. Özden C., Karadoğan N. (2024) Wheat Yield Prediction for Turkey Using Statistical Machine Learning and Deep Learning Methods. PAKISTAN JOURNAL OF AGRICULTURAL SCIENCES, Vol. 61(1), pp.1-7.

7. Fiorentini, M., Schillaci, C., Denora, M., Zenobi, S., Deligios, P., Orsini, R., Santilocchi, R., Perniola, M., Montanarella, L., & Ledda, L. (2024). A machine learning modeling framework for Triticum turgidum subsp. durum Desf. yield forecasting in Italy. Agronomy Journal, Vol. 116, pp. 1050–1070. [Online]. Available: https://doi.org/10.1002/agj2.21279.

8. Navid Mahdizadeh Gharakhanlou, Liliana Perez. (2024) From data to harvest: Leveraging ensemble machine learning for enhanced crop yield predictions across Canada amidst climate change. Science of The Total Environment. Vol. 951, p. 175764. [Online]. Available: https://doi.org/10.1016/j.scitotenv.2024.175764.

9. McKinney, W. (2018) Python for Data Analysis. O'Reilly Media.

10. Schierhorn, F., Hofmann, M., Gagalyuk, T. et al. (2021) Machine learning reveals complex effects of climatic means and weather extremes on wheat yields during different plant developmental stages. Climatic Change, Vol. 169(39). [Online]. Available: https://doi.org/10.1007/s10584-021-03272-0.

11. P. Hrytsiuk, T. Babych, S. Baranovsky, M. Havryliuk. (2023) Assessing of Climate Impact on Wheat Yield using Machine Learning Techniques. CEUR Workshop Proceedings, 3513, pp. 314–329.

12. Iqbal N, Shahzad MU, Sherif E-SM, Tariq MU, Rashid J, Le T-V, Ghani A. (2024) Analysis of Wheat-Yield Prediction Using Machine Learning Models under Climate Change Scenarios. Sustainability. Vol. 16(16), p. 6976. [Online]. Available: https://doi.org/10.3390/su16166976.

The article has been sent to the editors 12.03.25.
After processing 21.03.25.
Submitted for printing 30.03.25.